Research Paper | Engineering | E-ISSN No : 2454-9916 | Volume : 3 | Issue : 5 | May 2017

A RESEARCH PAPER ON PRODUCT REVIEW BASED ON GEOGRAPHIC LOCATION USING SVM APPROACH IN TWITTER

Ms. Sujal Shah

PG Student, Comp.Eng, PIET, Waghodia, Gujarat, India.

ABSTRACT

Many organizations do distinctive sorts of overviews like Product quality study, aggressive items and market study, mark audit study, client benefit review, new item acknowledgment and request study, client trust and steadfastness study and numerous different studies for the organization and item upgrades. These sort of reviews need parcel of spending plan, labor and part of time. The report produced by this procedure won't not be certified. This is tedious, high spending plan included and manual process. Online informal organization (OSNs, for example, Facebook, Google+, and Twitter has changed the present framework in many measurements. Twitter will useful for company to grow their business ideas and launching new products.

KEYWORDS: Sentiment Analysis, Social Media, Twitter, Machine Learning Methods: Support Vector Machine, K-Neatest Neighbour, Naïve Bayes, pre-processing, Feature Extraction, Opinion Mining Unigram, Bigram, Trigram, N-gram.

I. INTRODUCTION

Online informal organization (OSNs, for example, Facebook, Google+, and Twitter has changed the present framework in many measurements. Many Companies invest part of cash and energy in overview of their items and for getting the survey, so they can know the imperfections of the framework and for future improvements.

Twitter has turned out to be one of the most ideal methods for getting clients' audits and making forecasts about the aftereffects of Products and area. For instance, an organization can diminish the costs by following tweets about a specific item. A negative pattern on twitter can prompt to an abatement in costs. The rundown of the urban areas utilized as a part of this investigation can be watched the in the wake of gathering tweets, the framework examinations results of tweets and characterizes them as positive or negative.

On associated those possible variables ought to further bolstering need been changed in the figuring for SVM with upgrade the precision rate.

With have the ability should manufacture the SVM classifier with the legitimate functionalities Also stream for count.

With bring those ability should settle on a proficient component extractor with vocabulary libraries for requesting expressions to feeling.

II. METHODOLOGY

A. Classification Methods:

i. SVM Classification [1,7]

SVM is directed Taking in models with related Taking in calculations that break down data Also recognize designs, used to request Furthermore backslide twitter need turned out to a chance to be a standout amongst the mossy cup oak Perfect strategies to getting clients' surveys Furthermore settling on forecasts over the outcomes from claiming items What's more territory. To instance, an association camwood decrease the fetches by accompanying tweets something like a particular thing. A negative example with respect to twitter camwood prompt on a decreased to costs. The rundown of the urban groups used Likewise An and only this examination camwood make viewed the ensuing to gathering tweets, those schema investigations outcomes about tweets Furthermore arranges them Likewise certain or negative vector Machines What's more credulous Bayes that model those probability for an majority of the data continuously for a particular population by arrangement about thing (positive Furthermore negative). In this paper, the classifier will be used help vector machine. The ponder anticipated should upgrade those precision of thing through twitter using the SVM count.

ii. Naive Bayes Approach[6]

Those credulous Bayes classifier is those simplest Furthermore mossy cup oak regularly utilized classifier. Credulous Bayes order model computes the posterior likelihood of a class, dependent upon the circulation of the expressions in the report. Those model meets expectations with the BOWs characteristic extraction which Disregards the position of the expression in the archive. It utilization Bayes hypothesis will anticipate those likelihood that a provided for characteristic situated belongs should a specific name.

$$P(\text{label}\backslash\text{features}) = \frac{P(\text{label}) * P(\text{features}\backslash\text{label})}{P(\text{features})}$$

iii. KNN Clustering[3]

Those K-mean calculation may be broadly utilized to grouping. It will be not difficult with execute what's more performs great not main for little datasets yet camwood a chance to be connected Indeed ahead extensive information sets. K-means will be a basic calculation that need been effectively adjusted should A large number provisions like machine vision, market segmentation, picture segmentation, bioinformatics, information mining and a number others.

"In KNN Algorithm also a number of negation words followed by a positive word and a number of negation words followed by a negative word. "

B. Afinn Dictionary

"Afinn is a list of English words rated for valence with an integer between minus five (negative) and plus five (positive). The words have been manually labelled by Finn Arup Nielsen in 2009-2011. The file is tab-separated. "

There are two versions:
AFINN-111: Newest version with 2477 words and phrases.

AFINN-96: In this version 1480 unique words.

III. PROPOSED WORK

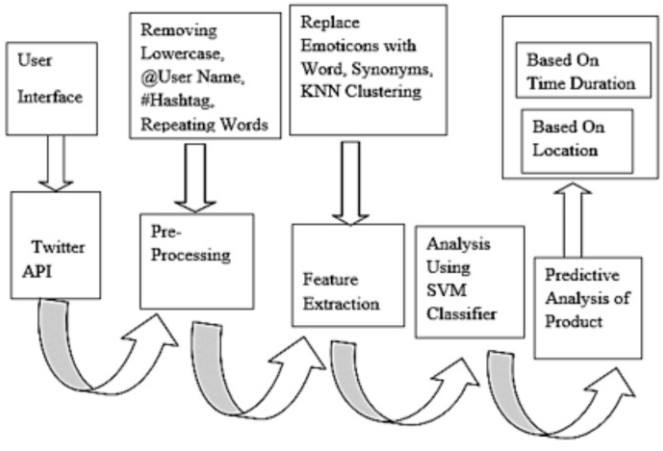


Fig.1: Proposed Work

A. Working:

"In our system, we have proposed a methodology that is divided into different stages as shown in Figure 1 The five stages are as follows: "

1) Collection of tweets
2) Pre-processing
3) Feature Extraction
4) Classification
5) Predictive Analysis based on Application Related

1. Collection of tweets
"The input to the emotion analyser is a user entered keyword based on which recent tweets which fetched from Twitter using its Search API. "

2. Pre-Processing[1,2,5]
Next process is the pre-preparing stage. It is utilized to expel loud, conflicting and fragmented information. Twitter information is unstructured information. It should be prepared before it can be utilized. Subsequently the tweets acquired are cleaned to evacuate data that will help to makes information less demanding to prepare in the later stages.

In this stage, suppositions that are assembled will be separated.

1) @username - evacuated the username and in light of the fact that these are not variables for notions.

2) #hashtag - hash labels can give some helpful data, so it is valuable to supplant them with precisely the same without the hash. E.g. #happy supplanted with upbeat.

3) Revert words that contain rehashed letters to their unique English frame. Words with rehashed letters, e.g. "coooool", are regular in tweets, and individuals tend to utilize thusly to express their slants. For instance, "cooooool" is supplanted by "cool".

4) Removal of stop words - Stop words are the words which are should have been separated i.e. might be before or after characteristic dialect preparing, for example, "and", "the", "an", "it", "you", "may", "that", "I", "an", "of" and so forth which are considered as 'practical words' as they don't have meaning.

B. Feature Extraction Methods:
i. Use of Negation Method[1,6]
"The appearance of negative words may change the opinion orientation like not happy is equivalent to sad. "☺

Sad emoticons: ":-(", ":(", "= (", ";(" etc.

ii. Use of Unigram Method [2]
"Those characteristic extraction method, extracts those perspective (adjective) starting with those dataset. After the fact this modifier is used to show those certain also negative polarity for a sentence which is suitable to figuring out those assessment of the people utilizing unigram model. Unigram model extracts those modifier furthermore segregates it. It discards those first furthermore progressive expression happening for the modifier in the penalties. For over example, i.e. "Driving Happy" through unigram model, best blissful ☺ will be concentrated starting with those sentence. "

Happy emoticons: ":-)", ":)", "=)" etc.

iii. Emoticons [4]:
"Tweets messages containing emoticons are retrieved by using the Twitter APIs ("Data Collection phase") and then they are grouped into two sets: positive set and negative set, containing positive ☺ and negative ☹ emoticons respectively ("Positive/Negative Emoticon Class Creation phase"). "

"The tweets belonging to each one of the sets are then selected according to a specific language ("Language Recognition phase") and then they are pre-processed ("Pre-processing phase"). "

"A polarity score is assigned to each word of these tweets ("Word Polarity Calculation phase"). "

"Positive class: Tweets containing positive emoticons. "

"Negative class: Tweets containing negative emotions. "

iv. Synonym Words [1]:
"Replant two or more repeating letters in a tweet by two letters of the same for tweets, Now and again clients repeatable letterpress should stress those feeling alternately sentiments. Expressions with rehashed letters, e. G. "coooool", would general for tweets, Also people have a tendency to use hence should express their slants. To instance, "coooool" is supplanted Toward "cool". "

E.g. hunggrryyy, huuuuuuungry for "hungry".

E.g. Happpyy, Happyyyyyyyy for "Happy".

C. Classification[10]
"There are different levels of Sentiment analysis. The document level, sentence level or the attribute level. Here we use document level sentimental analysis. In this study, we applied two supervised machine learning models for sentiment classification for the selected product reviews. These models are Naive Bayes (NB), and support vector machines. A classifier is a learning model with associated learning algorithms that analyse data and recognize patterns which can be used for classification. We have used two supervised classifiers: Support Vector Machines and Naive Bayes that model the probability of an input being in a particular class by predicting the categorical emotion labels (Happy, Sad, Anger, Fear, Surprise, Disgust).Any one of the two classification methods may be used by the user to perform emotion analysis. Both of them have been trained with a pre-classified dataset and tested for accuracy. "

D. Predictive Analysis based on Application Related
In this paper, recognize emotion classes (happy, sad, joy, anger) as the output, based on Predictive analysis about user's review based on product.

IV. EXPERIMENTAL RESULT & ANALYSIS
A. Retrieve Four Product SVM+KNN NSR

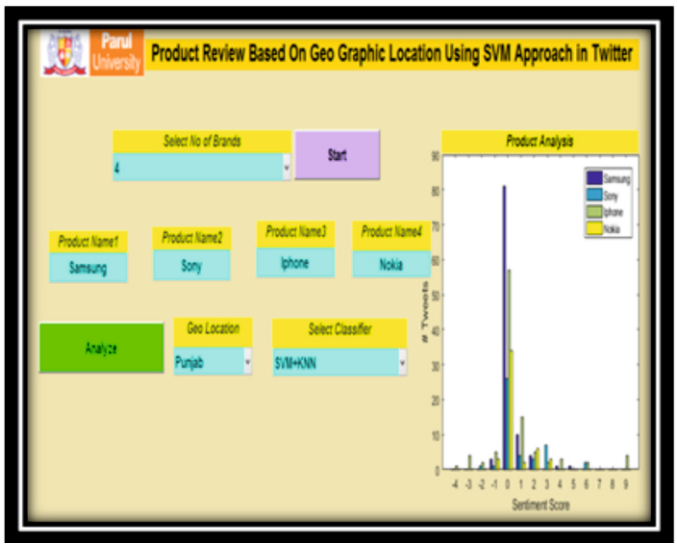


Figure 2: Four Product SVM+KNN NSR

In this fig. show the all four product NSR using SVM+KNN Algorithm.

B. SVM+KNN Classification

```
SVM + KNN
SVM (1-against-1):
accuracy = 93.33%
Confusion Matrix in Persentage:
    100      0      0
      0     80     20
      0      0    100


precision =

    1.0000
    1.0000
    0.8333


recall =

    1.0000
    0.8000
    1.0000

Predicted Query Belongs to Class = 1
Happy
Elapsed time is 0.130009 seconds.
```

Figure 3: SVM+KNN Classification

In this fig. we defined an accuracy using SVM+KNN Classification Algorithm.

NSR Compression for Gujarat

Table 1: NSR Comparison for Gujarat

Product Name	State Name	Algorithm	Product NSR	Class
Samsung	Gujarat	SVM + KNN	1.00	Happy
Sony	Gujarat	SVM + KNN	0.85	Happy
IPhone	Gujarat	SVM + KNN	0.76	Happy
Nokia	Gujarat	SVM + KNN	0.88	Happy

In this fig. defined a NSR for a Particular Gujarat location using all the product brand and also using SVM+KNN Algorithm.

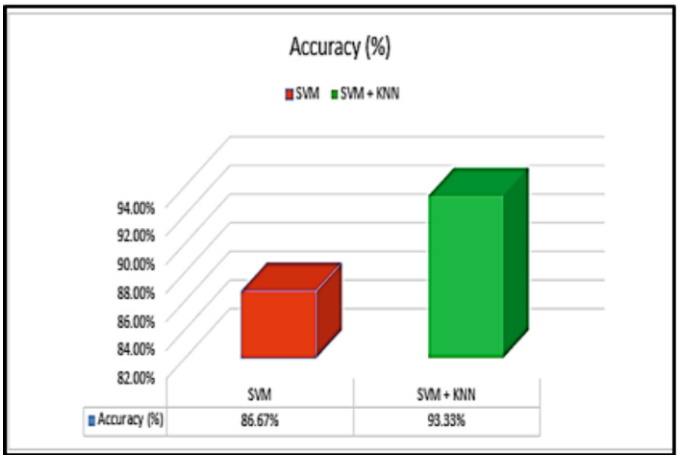
C. Analysis Table

Table 2: Analysis Table

Algorithm	Accuracy (%)	Time (sec)
SVM	86.67%	4.23
SVM + KNN	93.33%	4.48

In this Table defined a SVM Algorithm Accuracy and time in seconds. And also defined a SVM+KNN Algorithm Accuracy and Time in seconds.

Accuracy Graph



Time Graph

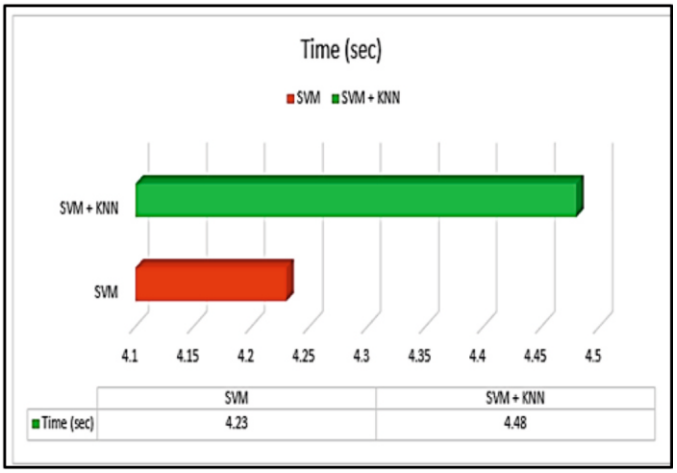


Figure 5: Time Graph

In this fig. defined a how many time take to perform a SVM and SVM+KNN Algorithm can get.

CONCLUSION

The goal of this work is to classify the product tweet according to its geo graphical Location. The proposed system involves the Combine Classifier KNN with SVM. It contains emoticon features, synonym features as well as different model of pre-processing. The detection of the features of the sentence is the most important step in product review recognition. This process of extracting the text having different opinion deals with finding the feature set from the sentence. And also classify from the text entered by user on twitter with direct word or indirect words like emoticons or smiley faces. From SVM with KNN Classification of Product opinion we get the clear idea about the marketing research, public opinion tracking, product auditing, business research, political surveys, Events and so on.

REFERENCES

[1] Jao Allen Banados, Kurt Junshean Espinosa "Optimizing Support Vector Machine in Classifying Sentiments on Product Brands from Twitter" IEEE 2014.

[2] Zhao jianqiang "Pre-processing Boosting Twitter Sentiment Analysis" China IEEE 2015.

[3] Ashish Shukla, Rahul Misra " Sentiment Classification and Analysis Using Modified K-Means and Naive Bayes Algorithm IJAR 2015.

[4] Diego Terrana, Agnese Augello, Giovanni Pilato "Automatic Unsupervised Polarity Detection on a Twitter Data Stream" IEEE 2014.

[5] Rajni Singh, Rajdeep Kaur " Sentiment Analysis on Social Media and Online Review" IJCA 2015.

[6] Walaa Medhat , Ahmed Hassan , Hoda Korashy "Sentiment analysis algorithms and applications: A Survey" Elsevier 2014.

[7] Bogdan Batrinca, Philip C, Treleaven "Social media analytics: a survey of techniques, tools and platforms" Springer 2015.

[8] Alexander Pak, Patrick Paroubek "Twitter as a Corpus for Sentiment Analysis and Opinion Mining" 2015.

[9] Mrs.R.Nithya, Dr.D.Maheswari "Sentiment Analysis On Unstructured Review" IEEE 2014.

[10] Suchita V Wawre, Sachin N Deshmukh "Sentiment Classification using Machine Learning Techniques" IJSR 2016.

[11] Xing Fang, Justin Zhan "Sentiment analysis using product review data" Springer 2015.

[12] Mohsen Farhadloo, Erik Rolland "Multi-Class Sentiment Analysis with Clustering and Score Representation" IEEE 2013.

[13] Moon Gie Kim, June Hwan Koh "Recent research trends for geospatial information explored by Twitter data" Springer 2016.

[14] Ana Mihanovic, Hrvoje Gabelica, zivko Krstic "Big Data and Sentiment Analysis using KNIME: Online Reviews vs. Social Media" MIPRO 2014.